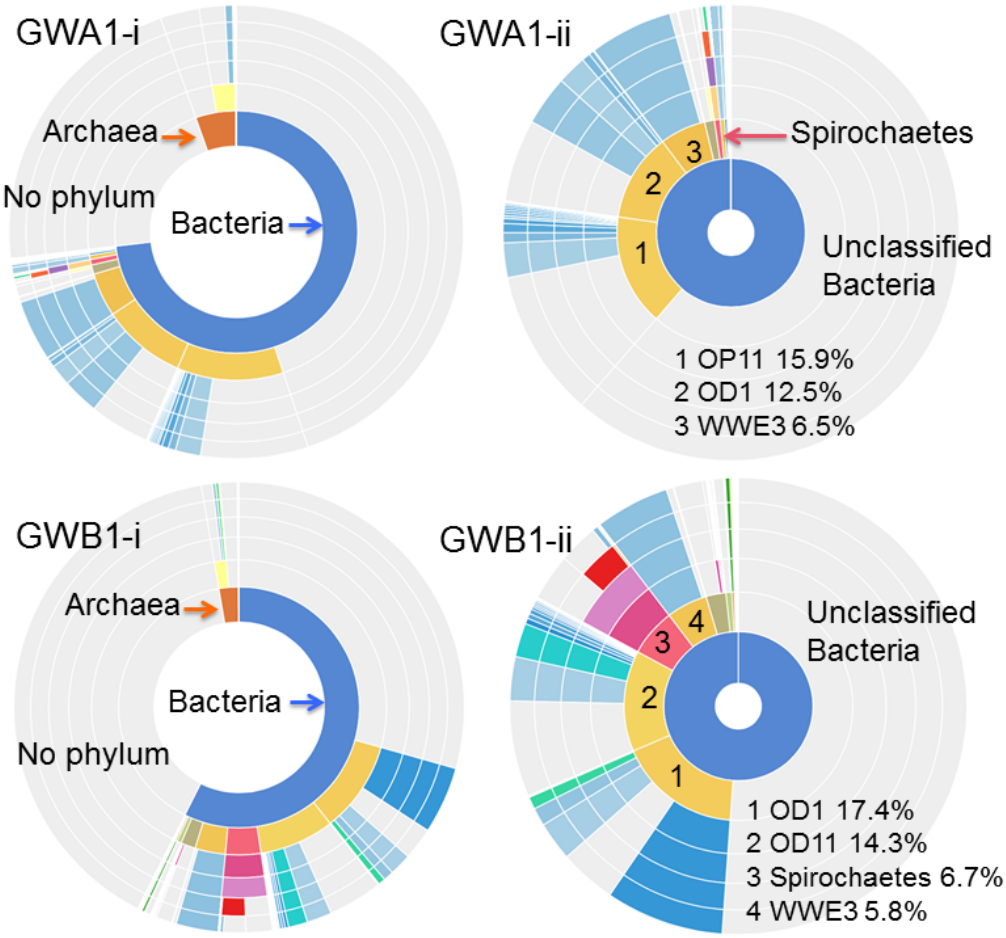
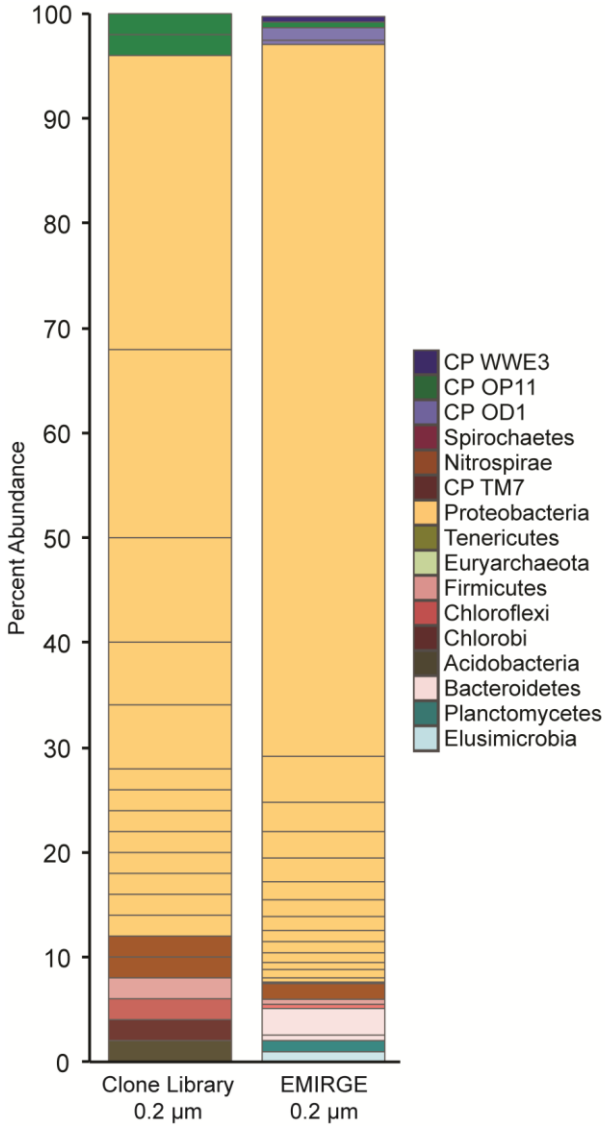
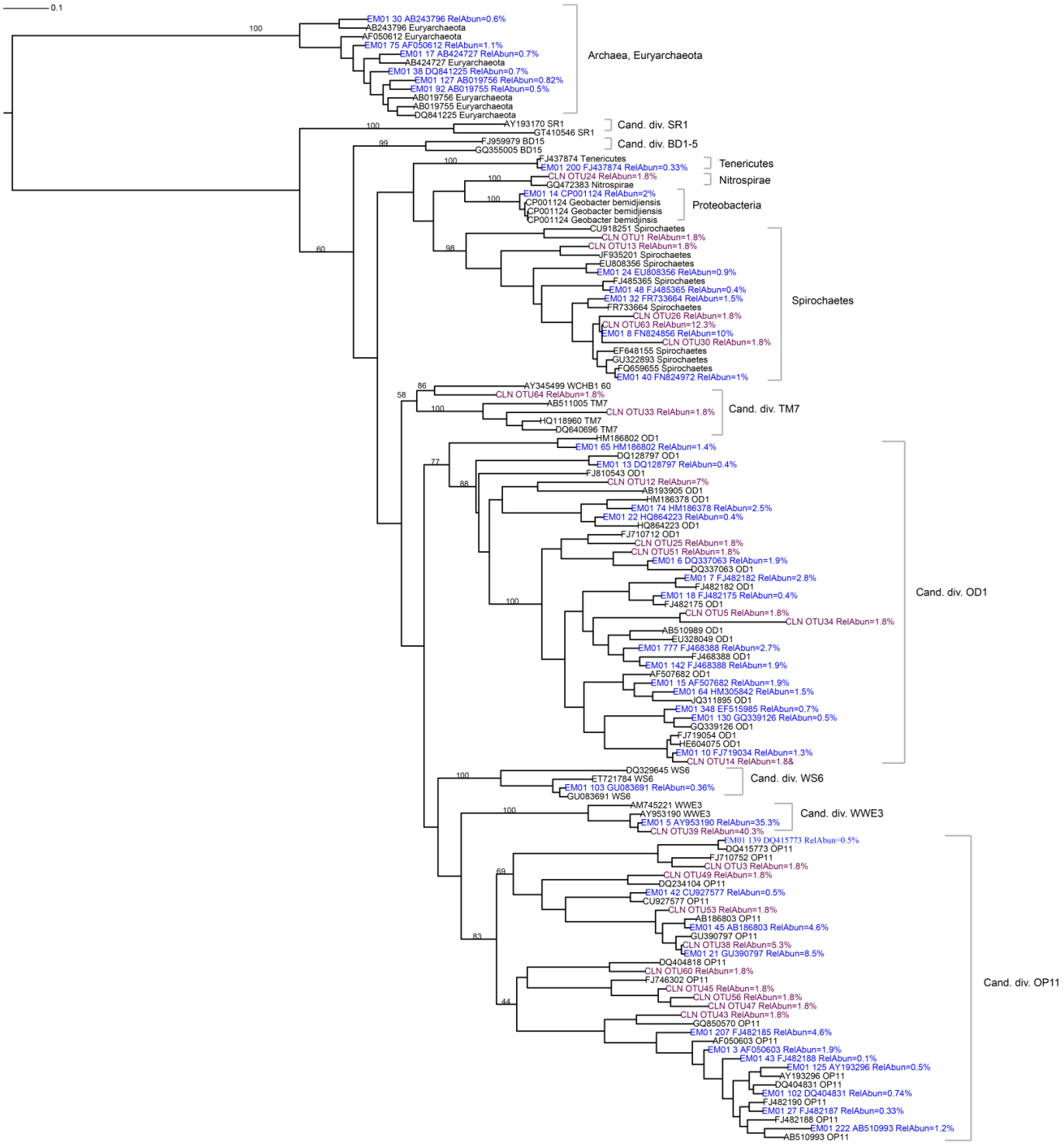


Supplementary Figure 1. Comprehensive overview of the microbial composition of the GWA1 (post 0.2 µm background groundwater) and GWB1 (post 0.2 µm acetate-amended groundwater) samples. The phylogenetic wheels report the fraction of the assembled genomic sequence (encoding ~250,000 genes for A1 and ~364,000 genes for B1) that could be classified at each taxonomic level, based on phylogenetic profiles of assembled fragments. Profiles were generated by comparison to sequenced genomes. Assignments required ≥ 50% of genes to have a best match at a given phylogenetic level. For example, if ≥ 50% of genes on a fragment had a best match to the same species, the classification would be to that species. If < 50% of genes had a best match to organisms from the same class but ≥ 50% matched to the same phylum, the classification of the fragment would be only established at the phylum level. In GWA1-i and GWB1-i the inner circle shows the fraction of the sample classified as Bacteria (blue), Archaea (orange) or unclassifiable at the Domain level (largely phage). Note the low signal from Archaea in GWB1. In GWA1-ii and GWB1-ii, classifications of just the Bacterial components (blue) are shown. Because there are only very

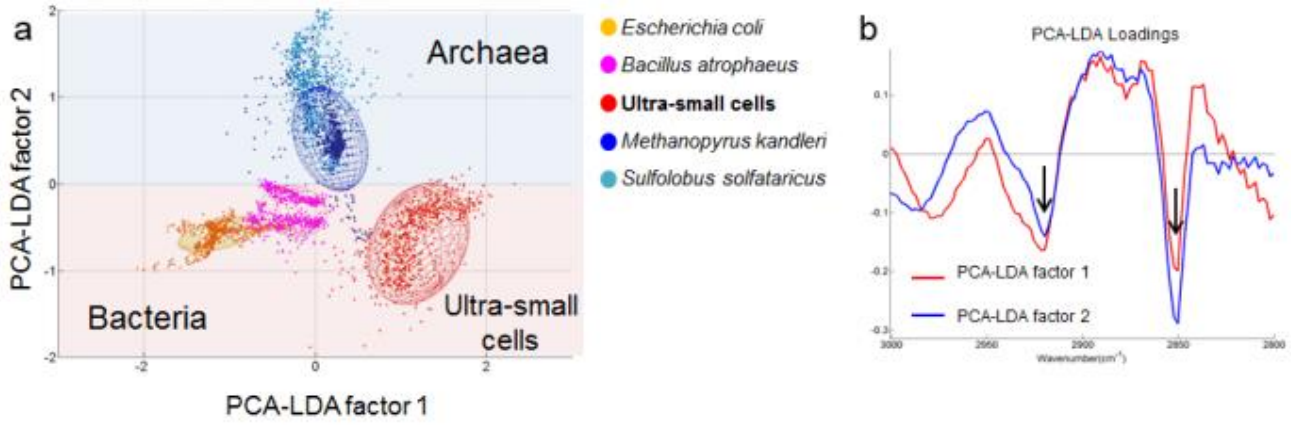
small numbers of reference genomes for OD1, OP11 and WWE3, most CP genome sequences are classified as ‘unclassified bacteria’. 140 CP radiation bacterial genomes have been reconstructed from GWB1 (including 68 OD1, 59 OP11 and 5 WWE3 genomes) and will be reported elsewhere. Note the low abundance of Bacteria from lineages other than the CP radiation.



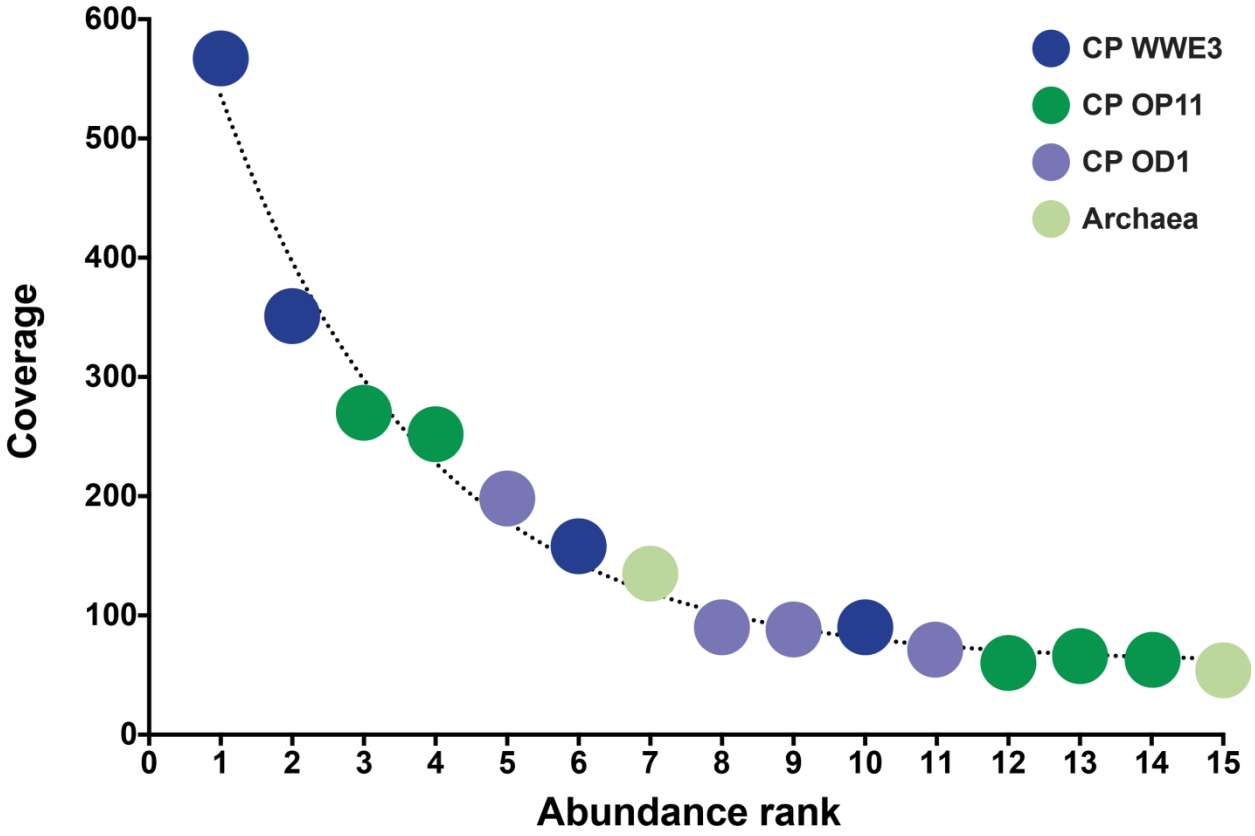
Supplementary Figure 2. Phylum level breakdown of the 16S rRNA gene sequences recovered from the 0.2 µm filter, based on EMIRGE and clone library information (clustered at 97 %). Abbreviation: CP, candidate phyla.



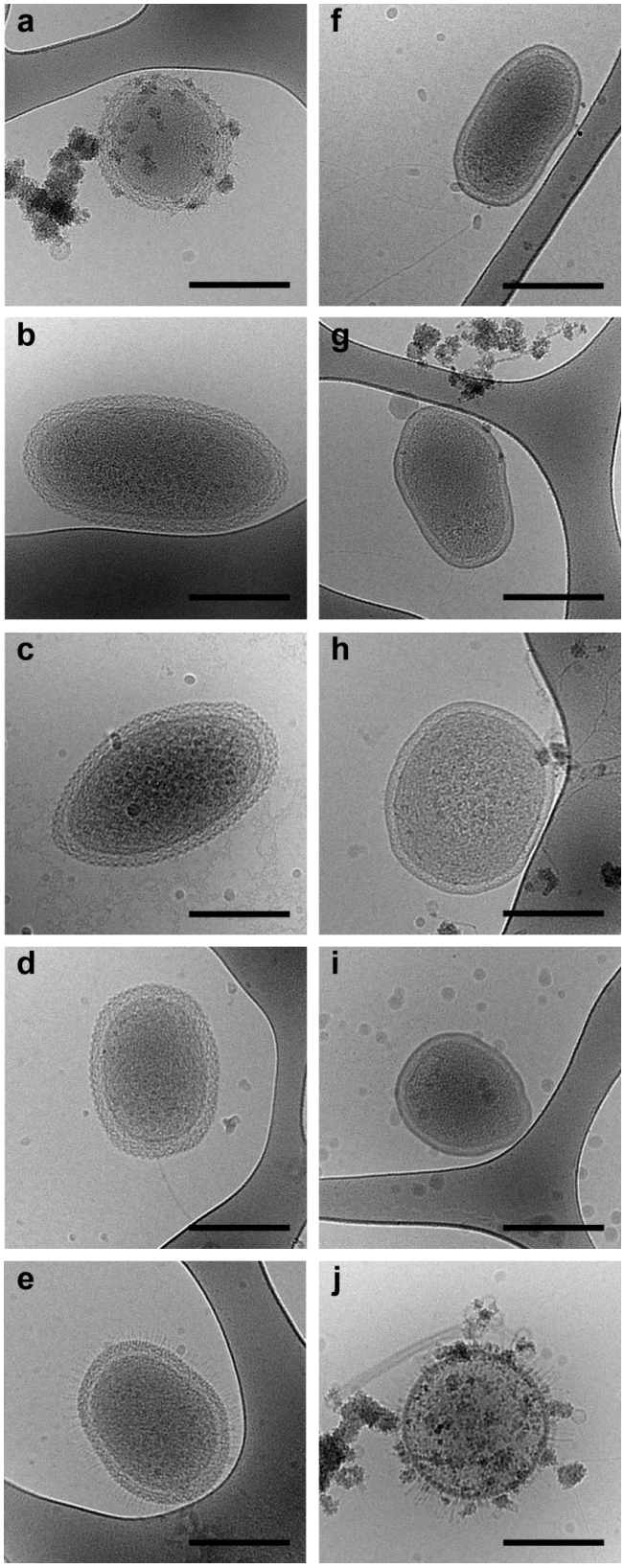
Supplementary Figure 3. Maximum likelihood tree showing the phylogenetic placement of the WWE3-OP11-OD1 phyla. To prepare the tree, all 16S rRNA gene sequences recovered from cells that passed through the 0.2 µm filter were combined with database sequences (representing the most closely related taxa) and aligned. Names in blue denote EMIRGE reconstructed sequences, while names in purple denote clone library OTUs. RelAbun = Relative abundance percentage in EMIRGE or clone library respectively.



Supplementary Figure 4. Spectra comparison of reference bacteria and archaea to ultra-small bacteria. Principal Component Analysis and Linear Discriminant Analysis (PCA-LDA) of the spectra in the CH vibration region (3000 - 2800 cm^{-1}). **a**) Two dimensional PCA-LDA score plots reveal an excellent separation of archaea and bacteria along the PCA-LDA factor 2. PCA-LDA factor 1 seems to be size dependent. Each ellipse covers an area of 95% confidence level. **b**) PCA-LDA factor 2 reveals two sharp negative spectral features at 2920 cm^{-1} and 2850 cm^{-1} (see arrows), which are associated with CH_2 bond symmetric and anti-symmetric vibrational stretching. Note that the spectra of the reference bacteria and archaea were from our validated database library.

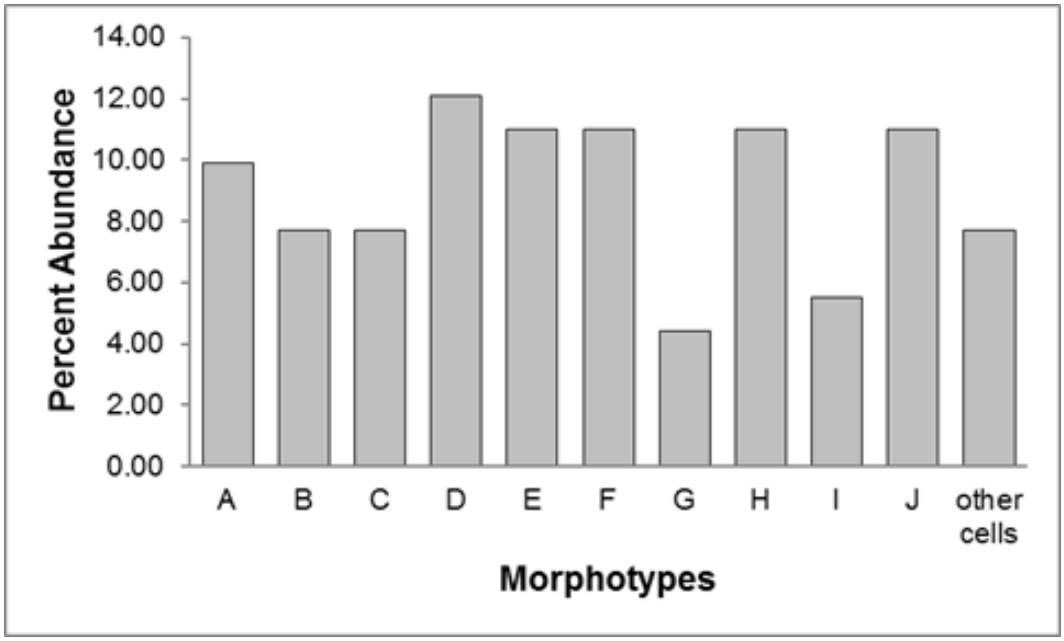


Supplementary Figure 5. Rank abundance curve illustrating dominance of the GWA1 microbial community (prior to acetate amendment) by WWE3 bacteria, with lower abundances of OP11, OD1 bacteria and archaea. Sequence coverage (y-axis) is directly related to organism abundance. Coverage values were computed by read mapping to contigs generated in two sub-assemblies. In some cases coverage values were only available from one sub-assembly.

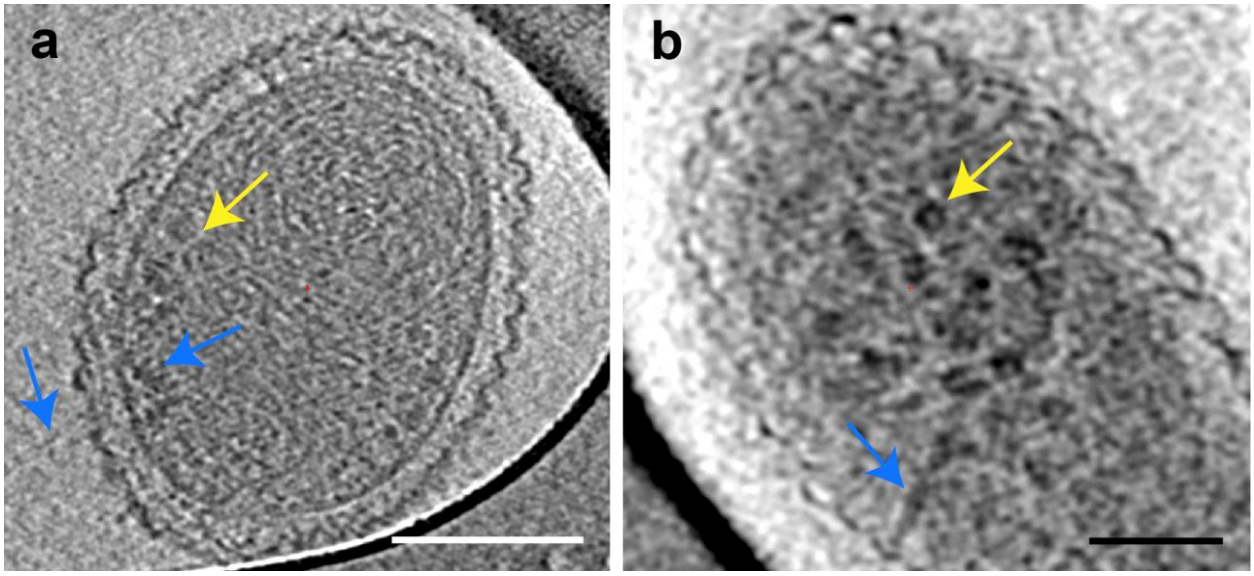


Supplementary Figure 6. Morphotypes of ultra-small cells. Cryo-transmission electron micrographs documenting size and morphological features of ultra-small bacterial cells (**a**-**j**). The cell walls have an intriguing architecture, with a remarkable and distinct S-layer (**a**-**e**).

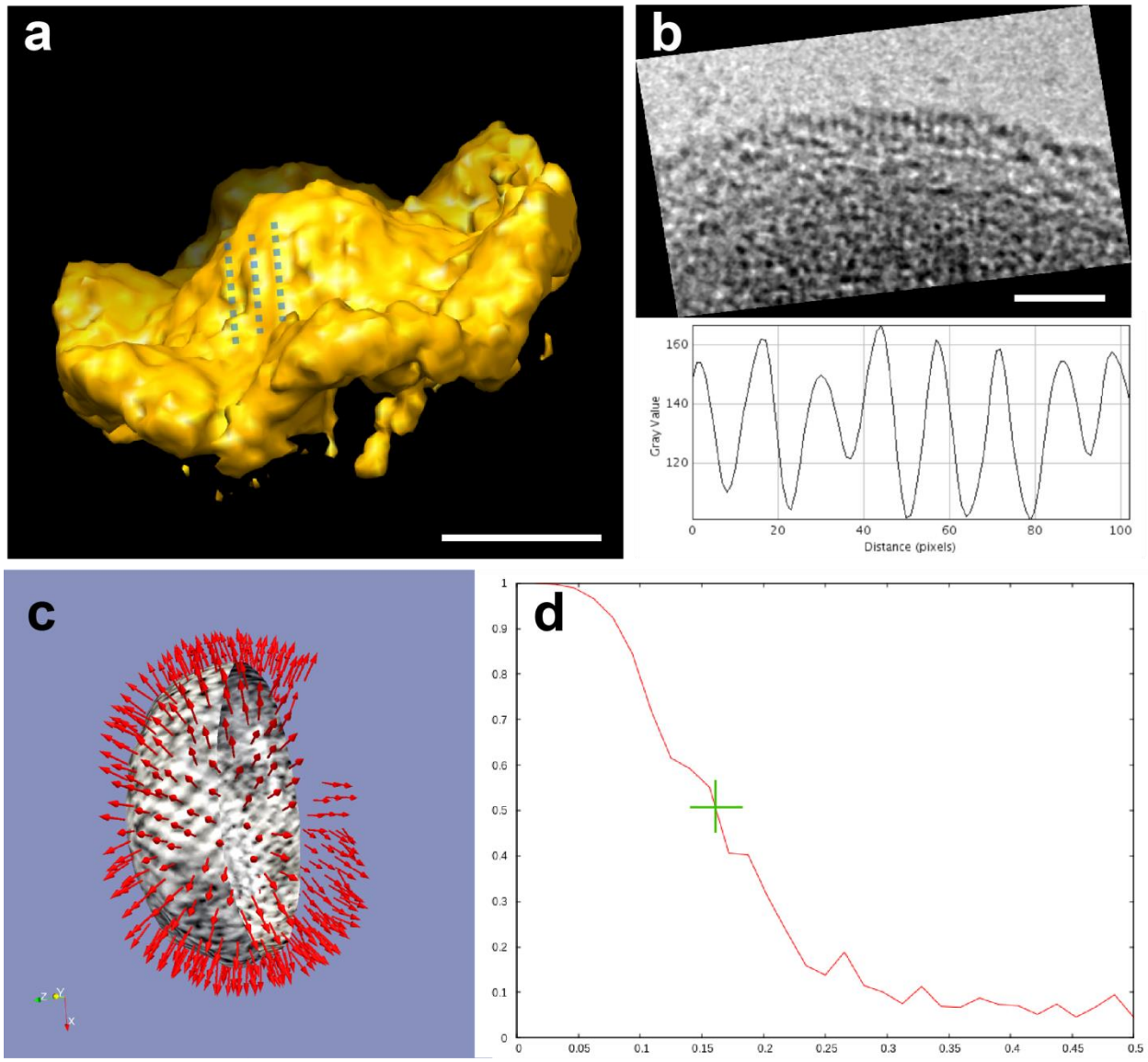
Numerous radiating pili-like structures cover the surfaces of cells in **a**), **e**) and **j**). Polar pili-like structures occur on cells **d**), **f**) and **g**). The cells in **a**) and **j**) have aggregates of nanoparticles attached to the outer membrane. One bacteriophage is associated with the surface of the cell in **g**). Seventy-eight 2D images and 13 tomograms were analysed to determine cell morphology. Scale bars are all 200 nm.



Supplementary Figure 7. Breakdown of the 10 most common morphotypes of ultra-small cells. Categories **A** to **J** refer to the different morphotypes of bacterial cells shown in **Supplementary Fig. 6** (see also **Supplementary Note 3**).

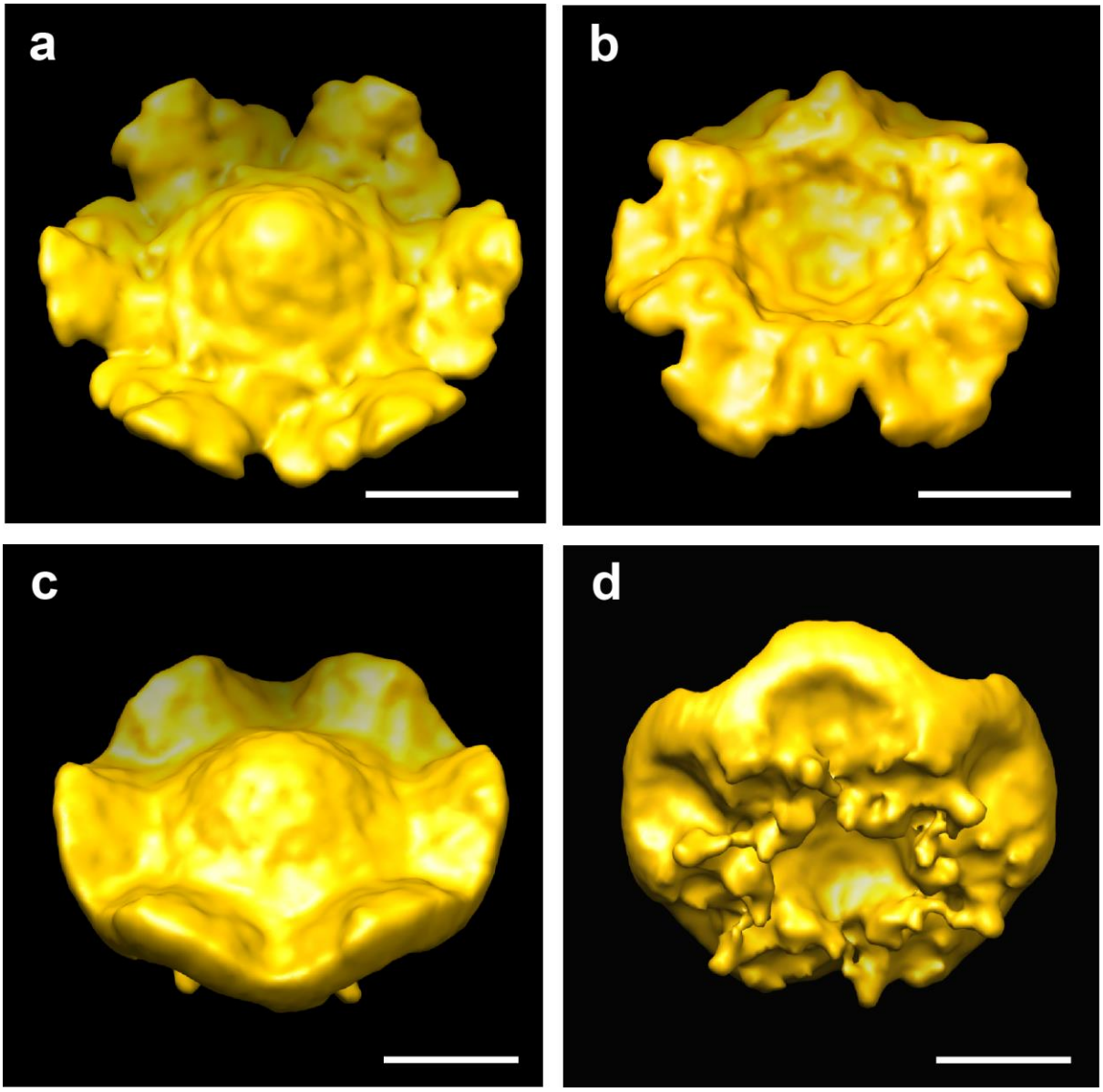


Supplementary Figure 8. Structural features of ultra-small cells in 3D slices. **a**) and **b**) Image of structures inferred to be ribosomes (yellow arrows). **a**) A pilus-like structure extends from outside of the cell, across the cell wall, and connects to an arc of high density just inside the inner membrane (blue arrow). For context, see **Supplementary Movie 2**. **b**) A long tubular density crossing the cytoplasm from the centre of the cell towards the cell wall (blue arrow). Scale bars are 50 nm.

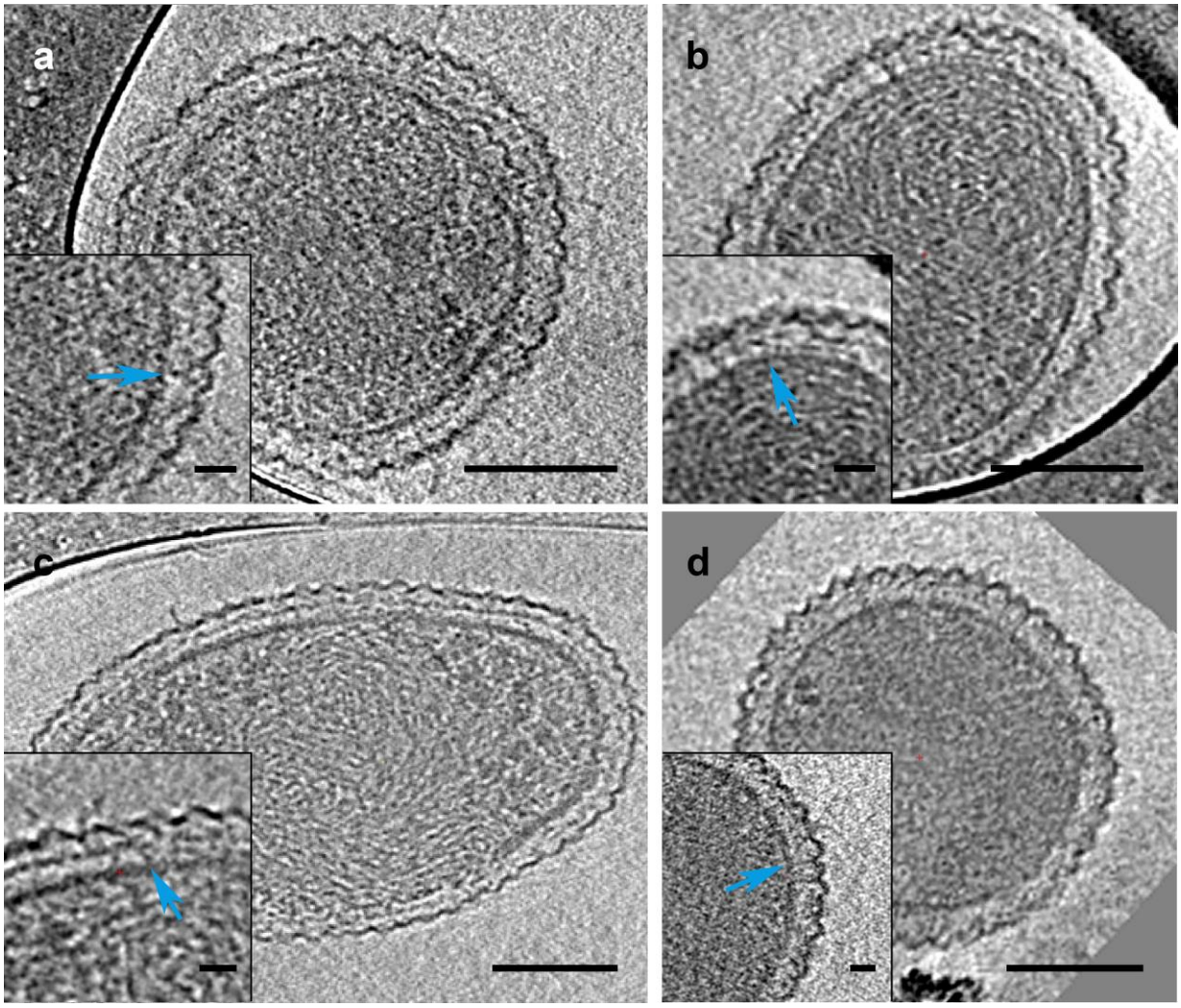


Supplementary Figure 9. Resolution of averaged subtomographic S-layer reconstruction. **a**) Side view of the un-symmetrized sub-volumetric averaged structure of the S-layer repeating unit. At this view ~1 nm in diameter globular densities are in-register (dotted blue lines). Scale bar is 10 nm. **b**) 2D projection of the boundary of a cell, in grey-scale, showing a pattern of lines embedded within the S-layer; the plot profile below the image displays grey-scale value versus pixel number. The peak-to-peak distance corresponds to ~3.3 nm for the parallel lines, matching exactly the spacing between the densities in **a**). Scale bar is 25 nm. **c**) Set of normal vectors to each sub-volume selected from one intact cell reconstruction used for the computation of the iterative alignment and averaging. Because the cells are near spherical or oblate, the angular coverage is good, even using a single data set. The geometry of the cells made the missing wedge problem easy to eliminate. **d**) Fourier shell correlation curve computed for the un-symmetrized reconstructions. A value of 0.5 corresponds to ~0.17

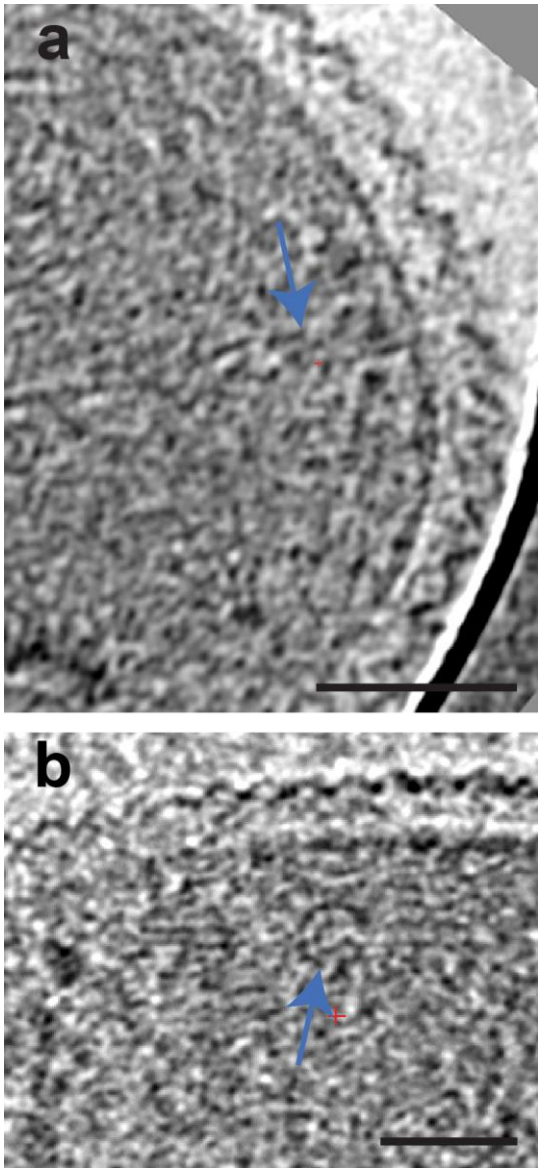
pixel^{-1} or 3.3 nm (0.56 nm per pixel). All images were acquired at approximately the same defocus value (~3.6 μm). The basic repeating motif is hexagonal and made of small subunits approximately 1 nm in diameter.



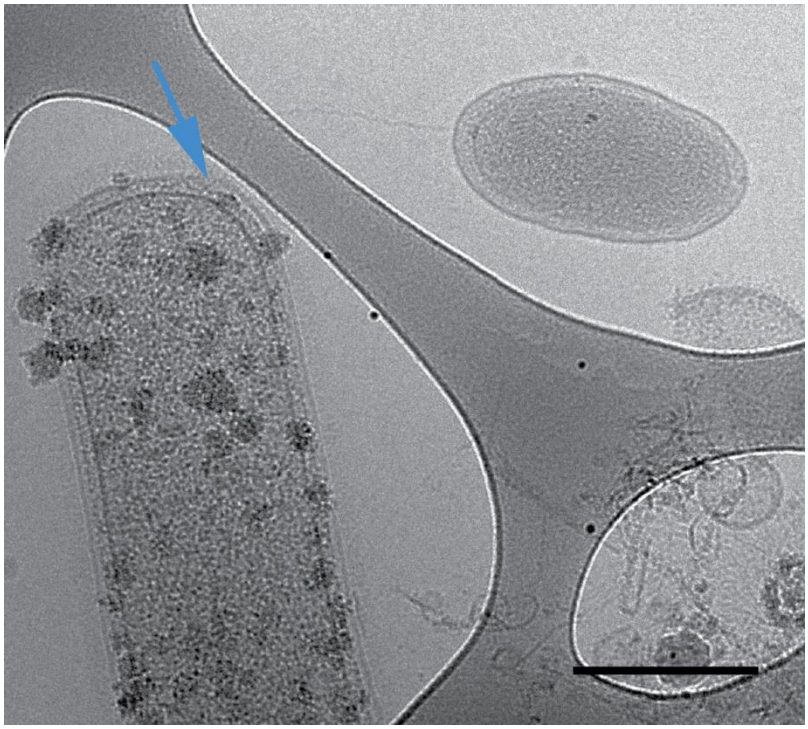
Supplementary Figure 10. Tomographic sub-volumetric averaged structure of the S-layer repeating unit, obtained from three cells. These structures are a subset of 785 units (See **Supplementary Note 5** for full details). **a**) and **b**) show top and bottom isosurface renderings of the symmetrized repeating unit of the S-layer, whereas **c**) and **d**) show top and bottom views of the same reconstruction rendered using higher histogram values, thus representing a larger enclosed mass than in **a**) and **b**). The top views show the exterior surface and bottom views show the subunit attached to the cell wall. Connectors to anchor the structure to the peptidoglycan can be clearly identified. Scale bars are 10 nm.



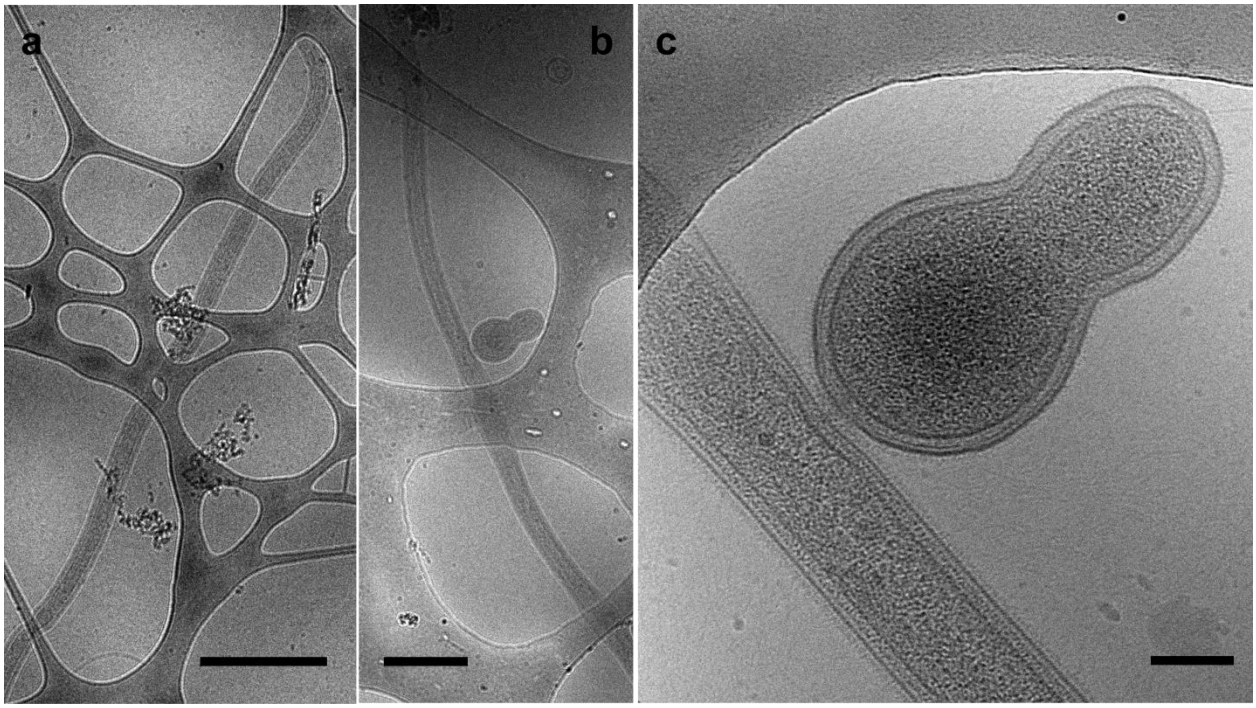
Supplementary Figure 11. Slices from 3D reconstructions of tomograms showing typical examples of ultra-small cells where the S-layers seem to be anchored through the peptidoglycan to the cytoplasmic membrane (e.g. blue arrows in the inserts). This is consistent with the sub-tomographic averaged structure presented in **Supplementary Fig. 10**. Scale bars are 100 nm. Scale bars of inserts are 20 nm.



Supplementary Figure 12. Slices from 3D reconstructions of tomograms showing unexplained features within the cytoplasm and cell walls of some ultra-small cells. **a**) A braided filament crossing the cell wall. **b**) A ring-shaped structure associated with a long filamentous structure. When the reconstruction is sliced at different angles, it is apparent that this ring surrounds the filament. Scale bars are 50 nm.



Supplementary Figure 13. Image of an ultra-small cell with a pilus-like structure that extends to the surface of another large cell with associated nanoparticle aggregates. Arrow points to pilus-like structure - cell connection. The larger cell is micron-sized. Scale bar 200 nm.



Supplementary Figure 14. Cryo-TEM projections of Spirochaete cells. **a)** and **b)** low resolution cryo-TEM images of Spirochaete cells (long filamentous cells crossing the field of view). Scale bars are 500 nm. **b**) shows low resolution image of **Figure 4c**. Low resolution cryo-TEM images were collected in low dose defocused diffraction mode. This imaging mode allows comprehensive surveys of microorganisms with minimum electron dose. **c**) Larger view image of a dividing ultra-small bacterium in contact with a Spirochaete cell shown in **Figure 4c**. Scale bar is 100 nm.

	Scaffold, gene number	Coverage	CLASSIFICATION
1	GWB1_contigs	770	WWE3_ACD25-ii
2	GWB1_scaffold_11715_13	283	OP11_ACD13_lineage
3	GWB1_scaffold_3136_18	250	OD1_ACD81_lineage
4	GWB1_scaffold_2354_10	204	OP11_ACD61_lineage
5	GWB1_contigs	185	OP11_ACD61_lineage
6	GWB1_scaffold_89_14	116	WWE3_ACD25
7	GWB1_scaffold_16588_10	112	OP11_ACD13_lineage
8	GWB1_scaffold_6378_18	103	OP11_ACD13_lineage
9	GWB1_scaffold_312_24	83	Spirochaetes
10	GWB1_scaffold_11151_9	78	OD1_novel_lineage_C
11	GWB1_scaffold_7361_9	65	OD1_ACD66,72,76_lineage
12	GWB1_scaffold_4661_8	50	Archaea
13	GWB1_scaffold_12214_9	44	OD1_RAAC4_lineage
14	GWB1_scaffold_2886_7	42	WWE3_lineage
15	GWB1_scaffold_260_107	39	OP11_ACD19_lineage
16	GWB1_scaffold_6915_4	38	OP11_ACD61_lineage
17	GWB1_scaffold_1026_58	36	OP11_ACD61_lineage
18	GWB1_scaffold_207_25	30	OP11_ACD30_lineage
19	GWB1_scaffold_57_242	29	OP11_ACD19_lineage
20	GWB1_scaffold_2498_13	29	Archaea
21	GWB1_scaffold_14987_5	27	OD1-i _lineage
22	GWB1_scaffold_1228_84	26	OP11_novel_lineage_A
23	GWB1_scaffold_638_22	24	OD1_RAAC4_lineage
24	GWB1_scaffold_3751_7	23	Archaea
25	GWB1_scaffold_372_47	22	OP11_ACD19_lineage
26	GWB1_scaffold_308_115	21	OD1_RAAC4_lineage
27	GWB1_scaffold_15218_12	21	Archaea
28	GWB1_scaffold_6290_5	20	OP11_ACD13_lineage
29	GWB1_scaffold_2265_9	20	OP11_ACD50_lineage
30	GWB1_scaffold_18725_9	20	OD1-i _lineage
31	GWB1_scaffold_3613_73	20	OD1_RAAC4_lineage
32	GWB1_scaffold_5775_2	20	OD1_ACD81_lineage
33	GWB1_scaffold_44_21	19	OP11_novel_lineage_A
34	GWB1_scaffold_2351_11	19	OP11_ACD50_lineage
35	GWB1_scaffold_5651_14	19	OD1_RAAC4_lineage
36	GWB1_scaffold_321_25	19	OD1_RAAC4_lineage
37	GWB1_scaffold_3446_9	19	OD1_RAAC4_lineage
38	GWB1_scaffold_1062_30	18	Archaea
39	GWB1_scaffold_11414_19	18	Archaea
40	GWB1_scaffold_346_164	18	OP11_ACD27_lineage
41	GWB1_scaffold_16593_8	18	OD1_RAAC4_lineage
42	GWB1_scaffold_42296_7	17	OD1_RAAC4_lineage

43	GWB1_scaffold_3209_6	17	OD1_novel_lineage_C
44	GWB1_scaffold_2699_13	15	Spirochaetes
45	GWB1_scaffold_7_114	15	OP11_ACD27_lineage
46	GWB1_scaffold_25440_9	15	OD1-i _lineage
47	GWB1_scaffold_1015_49	14	Archaea
48	GWB1_scaffold_11332_5	14	OD1-i _lineage
49	GWB1_scaffold_153_61	14	OD1_RAAC4_lineage
50	GWB1_scaffold_124_52	14	OP11_ACD30_lineage
51	GWB1_scaffold_3014_17	14	Archaea
52	GWB1_scaffold_4873_15	14	OD1_ACD81_lineage
53	GWB1_scaffold_15240_5	14	OD1_RAAC4_lineage
54	GWB1_scaffold_0_520	14	OP11_ACD13_lineage
55	GWB1_scaffold_558_52	13	Novel_group_B
56	GWB1_scaffold_2348_38	13	OD1-i _lineage
57	GWB1_scaffold_939_8	13	OP11_ACD13_lineage
58	GWB1_scaffold_2352_7	13	Archaea
59	GWB1_scaffold_6_130	13	OD1_novel_lineage_B
60	GWB1_scaffold_3351_28	12	OP11_ACD13_lineage
61	GWB1_scaffold_706_17	12	WWE3_lineage
62	GWB1_scaffold_301_186	12	OD1_RAAC4_lineage
63	GWB1_scaffold_192_114	12	OD1_ACD81_lineage
64	GWB1_scaffold_25_196	12	OD1_RAAC4_lineage
65	GWB1_scaffold_1816_10	12	OP11_ACD48_lineage
66	GWB1_scaffold_6679_12	12	OP11_novel_lineage_A
67	GWB1_scaffold_27929_8	12	OD1_RAAC4_lineage
68	GWB1_scaffold_270_53	12	OP11_ACD13_lineage
69	GWB1_scaffold_337_74	11	OP11_ACD48_lineage
70	GWB1_scaffold_9842_6	11	OP11_ACD13_lineage
71	GWB1_scaffold_13446_11	11	Archaea
72	GWB1_scaffold_20_78	11	OP11_ACD13_lineage
73	GWB1_scaffold_1757_10	11	OD1_novel_lineage_C
74	GWB1_scaffold_338_24	11	OP11_ACD13_lineage
75	GWB1_scaffold_1793_25	11	OP11_ACD13_lineage
76	GWB1_scaffold_9_603	11	OD1_ACD81_lineage
77	GWB1_scaffold_37017_8	11	OD1_RAAC4_lineage
78	GWB1_scaffold_10466_9	10	WWE3_lineage
79	GWB1_scaffold_17840_4	10	Novel_group_A
80	GWB1_scaffold_1076_27	10	OD1-i _lineage
81	GWB1_scaffold_13115_15	10	OP11_ACD13_lineage
82	GWB1_scaffold_2233_22	10	OD1_novel_lineage_A
83	GWB1_scaffold_35288_8	10	OD1_RAAC4_lineage
84	GWB1_scaffold_10073_16	10	OD1_ACD81_lineage
85	GWB1_scaffold_2152_37	10	Archaea

86	GWB1_scaffold_3799_25	10	OD1_RAAC4_lineage
87	GWB1_scaffold_75_14	10	OD1_ACD81_lineage
88	GWB1_scaffold_339_15	9	OD1_ACD81_lineage
89	GWB1_scaffold_2_35	9	Novel_group_A
90	GWB1_scaffold_3698_17	9	OP11_ACD50_lineage
91	GWB1_scaffold_5907_12	9	OP11_novel_lineage_A
92	GWB1_scaffold_2964_14	9	OD1_RAAC4_lineage
93	GWB1_scaffold_120_38	9	Archaea
94	GWB1_scaffold_3790_19	9	OP11_ACD30_lineage
95	GWB1_scaffold_146_75	9	Delta_novel
96	GWB1_scaffold_17077_2	9	OD1_novel_lineage_B
97	GWB1_scaffold_12649_1	9	Archaea
98	GWB1_scaffold_526_35	9	Archaea
99	GWB1_scaffold_323_59,60	8	Novel_group_F
100	GWB1_scaffold_581_32	8	OP11_ACD19_lineage
101	GWB1_scaffold_368_32	8	OD1_ACD81_lineage
102	GWB1_scaffold_40_37	8	OP11_ACD13_lineage
103	GWB1_scaffold_12719_8	8	Geobacter
104	GWB1_scaffold_14790_9	8	OD1-i _lineage
105	GWB1_scaffold_969_17	8	OD1_ACD81_lineage
106	GWB1_scaffold_1493_11	8	OD1-i _lineage
107	GWB1_scaffold_582_15	8	OP11_novel_lineage_A
108	GWB1_scaffold_15_37	8	OP11_ACD27_lineage
109	GWB1_scaffold_27352_10	8	OD1_RAAC4_lineage
110	GWB1_scaffold_2401_13	8	OD1_RAAC4_lineage
111	GWB1_scaffold_4433_14	8	OD1_RAAC4_lineage
112	GWB1_scaffold_657_28	8	OP11_ACD61_lineage
113	GWB1_scaffold_296_63	8	OD1_novel_lineage_C
114	GWB1_scaffold_802_38	8	OD1_ACD81_lineage
115	GWB1_scaffold_425_47	8	OP11_novel_lineage_C
116	GWB1_scaffold_810_18	8	OP11_ACD13_lineage
117	GWB1_scaffold_1771_21	8	OP11_ACD37
118	GWB1_scaffold_745_28	8	OD1_RAAC4_lineage
119	GWB1_scaffold_2476_12	8	OD1_RAAC4_lineage
120	GWB1_scaffold_5219_3	8	OD1_RAAC4_lineage
121	GWB1_scaffold_8704_17	8	OD1_ACD81_lineage
122	GWB1_scaffold_10118_10	8	OP11_ACD61_lineage
123	GWB1_scaffold_489_91	8	OD1_ACD81_lineage
124	GWB1_scaffold_129_81	8	Archaea
125	GWB1_scaffold_2950_22	8	Spirochaetes
126	GWB1_scaffold_479_43	8	OD1_ACD66
127	GWB1_scaffold_46_59	7	Archaea
128	GWB1_scaffold_867_60	7	OD1_RAAC4_lineage

129	GWB1_scaffold_4492_5	7	OD1-i _lineage
130	GWB1_scaffold_2866_13	7	OP11_ACD61_lineage
131	GWB1_scaffold_137_48	7	Archaea
132	GWB1_scaffold_605_47	7	OP11_ACD27
133	GWB1_scaffold_1229_75	7	OD1_RAAC4_lineage
134	GWB1_scaffold_3185_20	7	OD1_ACD81_lineage
135	GWB1_scaffold_12210_7	7	OD1_novel_lineage_A
136	GWB1_scaffold_645_29	7	Archaea
137	GWB1_scaffold_1951_1	7	OP11_ACD48_lineage
138	GWB1_scaffold_4063_26	7	OD1_ACD81_lineage
139	GWB1_scaffold_10449_2	7	OD1_novel_lineage_B
140	GWB1_scaffold_9238_2	7	OD1_RAAC4_lineage
141	GWB1_scaffold_3315_17	7	OD1_RAAC4_lineage
142	GWB1_scaffold_1102_22	7	OP11_novel_lineage_A
143	GWB1_scaffold_13057_6	7	OD1-i _lineage
144	GWB1_scaffold_2096_25	7	OP11_ACD27_lineage
145	GWB1_scaffold_4344_17	7	OD1_RAAC4_lineage
146	GWB1_scaffold_79_66	7	OP11_ACD19_lineage
147	GWB1_scaffold_21748_3	7	Archaea
148	GWB1_scaffold_7747_3	7	OP11_ACD13_lineage
149	GWB1_scaffold_1388_33	7	Spirochaetes
150	GWB1_scaffold_1952_37	7	Novel_group_D
151	GWB1_scaffold_310_52	7	OD1_RAAC4_lineage
152	GWB1_scaffold_3112_17	7	OD1_RAAC4_lineage
153	GWB1_scaffold_2009_20	6	OD1_ACD81_lineage
154	GWB1_scaffold_3961_18	6	OD1_RAAC4_lineage
155	GWB1_scaffold_3150_45	6	Archaea
156	GWB1_scaffold_2774_17	6	Archaea
157	GWB1_scaffold_6763_16	6	OD1_ACD81_lineage
158	GWB1_scaffold_1400_24	6	OD1_novel_lineage_B
159	GWB1_scaffold_436_27	6	Delta_novel
160	GWB1_scaffold_2491_12	6	Archaea
161	GWB1_scaffold_622_59	6	WWE3_lineage
162	GWB1_scaffold_1644_1	6	Spirochaetes
163	GWB1_scaffold_957_10	6	WWE3_lineage
164	GWB1_scaffold_5747_11	6	OP11_ACD48_lineage
165	GWB1_scaffold_11461_14	6	OD1_RAAC4_lineage
166	GWB1_scaffold_3074_14	6	OD1_novel_lineage_B
167	GWB1_scaffold_10606_6	6	OP11_ACD30_lineage
168	GWB1_scaffold_7976_9	6	Archaea(ARMAN)
169	GWB1_scaffold_4894_6	6	Archaea
170	GWB1_scaffold_3433_12	6	OP11_ACD13_lineage
171	GWB1_scaffold_2034_17	6	Archaea

172	GWB1_scaffold_8829_4	6	OP11_ACD27_lineage
173	GWB1_scaffold_792_36	6	Archaea
174	GWB1_scaffold_9338_4	6	OD1_RAAC4_lineage
175	GWB1_scaffold_4141_23	6	OD1-i _lineage
176	GWB1_scaffold_10246_10	6	OP11_ACD61_lineage
177	GWB1_scaffold_15086_8	6	OD1_RAAC4_lineage
178	GWB1_scaffold_3522_24	5	OP11_novel_lineage_D
179	GWB1_scaffold_5672_11	5	Novel_group_E
180	GWB1_scaffold_9202_2	5	Spirochaetes
181	GWB1_scaffold_9678_4	5	OD1_RAAC4_lineage
182	GWB1_scaffold_18824_21	5	OP11_ACD50_lineage
183	GWB1_scaffold_11948_3	5	Archaea
184	GWB1_scaffold_8024_14	5	OP11_ACD30_lineage
185	GWB1_scaffold_3148_12	5	OD1-i _lineage
186	GWB1_scaffold_545_28	5	Archaea
187	GWB1_scaffold_709_11	5	ACD58_lineage
188	GWB1_scaffold_4194_18	5	OD1-i _lineage
189	GWB1_scaffold_3342_18	5	OD1_RAAC4_lineage
190	GWB1_scaffold_4326_21	5	OD1_novel_lineage_C
191	GWB1_scaffold_11126_18	5	Archaea
192	GWB1_scaffold_13468_7	5	OP11_ACD48_lineage
193	GWB1_scaffold_14511_6	5	OD1_RAAC4_lineage
194	GWB1_scaffold_3308_15	5	Archaea
195	GWB1_scaffold_949_17	5	OP11_ACD30
196	GWB1_scaffold_4524_11	5	OD1_RAAC4_lineage
197	GWB1_scaffold_7073_10	5	Archaea
198	GWB1_scaffold_14853_9	5	OP11_ACD19_lineage
199	GWB1_scaffold_14654_2	5	OP11_ACD13_lineage
200	GWB1_scaffold_1142_27	5	Archaea
201	GWB1_scaffold_4200_21	5	OP11_ACD13_lineage
202	GWB1_scaffold_13450_8	5	OD1_RAAC4_lineage
203	GWB1_scaffold_7619_16	5	Archaea
204	GWB1_scaffold_12652_10	5	Archaea
205	GWB1_scaffold_14500_5	5	OD1_RAAC4_lineage
206	GWB1_scaffold_1882_17	5	OP11_ACD13_lineage
207	GWB1_scaffold_2788_3	5	Archaea
208	GWB1_scaffold_5057_9	5	OD1_RAAC4_lineage
209	GWB1_scaffold_11934_8	5	Archaea
210	GWB1_scaffold_25007_9	5	Archaea
211	GWB1_scaffold_30554_7	5	Archaea
212	GWB1_scaffold_6074_6	5	OP11_ACD30_lineage
213	GWB1_scaffold_12255_6	5	Archaea
214	GWB1_scaffold_14832_16	5	OP11_ACD13_lineage

215	GWB1_scaffold_6940_16	5	OP11_ACD30_lineage
216	GWB1_scaffold_7042_13	5	Archaea
217	GWB1_scaffold_6755_29	5	TM7_lineage
218	GWB1_scaffold_4739_9	5	OD1_RAAC4_lineage
219	GWB1_scaffold_9968_9	5	OP11_ACD30_lineage
220	GWB1_scaffold_6190_19	5	OD1_RAAC4_lineage
221	GWB1_scaffold_3630_15	5	OD1_RAAC4_lineage
222	GWB1_scaffold_13079_6	5	OD1_novel_lineage_B
223	GWB1_scaffold_5589_13	5	Archaea
224	GWB1_scaffold_16295_10	5	OD1_RAAC4_lineage
225	GWB1_scaffold_5572_12	5	Archaea
226	GWB1_scaffold_1729_10	5	WWE3_lineage
227	GWB1_scaffold_5370_7	5	Archaea
228	GWB1_scaffold_6585_26	5	Alpha_ACD16-related
229	GWB1_scaffold_17398_8	5	Spirochaetes
230	GWB1_scaffold_2507_17	5	OD1_novel_lineage_C
231	GWB1_scaffold_2990_4	5	OP11_ACD50_lineage
232	GWB1_scaffold_14534_1	5	OP11_ACD48_lineage
233	GWB1_scaffold_11146_9	4	Archaea
234	GWB1_scaffold_12185_8	4	OD1_RAAC4_lineage
235	GWB1_scaffold_4445_39	4	PER_lineage
236	GWB1_scaffold_17280_9	4	Archaea
237	GWB1_scaffold_4040_1	4	OP11_ACD13_lineage
238	GWB1_scaffold_21276_9	4	OP11_ACD13_lineage
239	GWB1_scaffold_11816_2	4	Novel_group_C
240	GWB1_scaffold_2431_15	4	OD1_RAAC4_lineage
241	GWB1_scaffold_9940_6	4	Archaea
242	GWB1_scaffold_17584_8	4	OD1_novel_lineage_A
243	GWB1_scaffold_3609_7	4	Archaea
244	GWB1_scaffold_7577_4	4	OP11_ACD50_lineage
245	GWB1_scaffold_11450_2	4	OD1_novel_lineage_C
246	GWB1_scaffold_13078_3	4	Archaea
247	GWB1_scaffold_7195_9	4	OD1-i _lineage
248	GWB1_scaffold_38852_2	4	OP11_novel_lineage_A
249	GWB1_scaffold_20849_2	4	OD1_RAAC4_lineage
250	GWB1_scaffold_23115_10	4	OP11_ACD50_lineage
251	GWB1_scaffold_8074_10	4	Archaea
252	GWB1_scaffold_7909_12	4	OD1_novel_lineage_B
253	GWB1_scaffold_22094_8	4	WWE3_lineage
254	GWB1_scaffold_10693_3	4	OP11_ACD48_lineage
255	GWB1_scaffold_10504_8	4	Archaea
256	GWB1_scaffold_9525_6	4	OP11_ACD13_lineage
257	GWB1_scaffold_16871_4	4	OD1_ACD81_lineage

258	GWB1_scaffold_12738_10	4	OP11_novel_lineage_B
259	GWB1_scaffold_33338_2	3	OP11_ACD13_lineage
260	GWB1_scaffold_55184_7	3	OP11_novel_lineage_C
261	GWB1_scaffold_55184_10	3	OP11_ACD48_lineage
262	GWB1_scaffold_5926_12	3	OD1_ACD81_lineage
263	GWB1_scaffold_59541_9	3	OP11_ACD13_lineage

Supplementary Table 1. Classification of community composition based on phylogenetic analysis of the ribosomal S3 proteins on assembled scaffolds > 5 kb (4.7 Gb of assembled sequence). Because genome fragmentation precluded identification of this protein for some organisms, we conducted three additional assemblies targeting highly sampled populations to confirm the representation of abundant organisms in the sample. Assignment of genome fragments to bins relied in part on similarity to a database of genome sequences for Candidate Phylum bacteria reported by Wrighton *et al.*[5]. For four populations, partial and full-length 16S rRNA gene sequences extracted from scaffolds also encoding ribosomal protein S3 genes were used to augment organism identification. For more complete and accurate representation of the most abundant populations see the results from sub-assemblies in **Figure 2**, **Supplementary Fig. 5** and **Supplementary Tables 3** and **5**.

	non-bacteria (pixels)	bacteria (pixels)	non-bacteria (%)	bacteria (%)
Sample1-Area 7c	0	238	0	100
Sample1-Area 7a-screen	0	24	0	100
Sample1-Area 7-screen	0	42	0	100
Sample1-Area8a-screen	1	39	2.5	97.5
Sample2-Area 6 cut	52	404	11.4	88.6
Total Pixels/Spectra: 800	53	747		
			2.8	**97.2**
Standard deviation			4.4	4.4

Supplementary Table 2. Determination of community composition (bacteria versus archaea) of ultra-small cells by Synchrotron infrared spectromicroscopy (with a 2 μm × 2 μm pixel resolution). Ultra-small cells on cryo-TEM grids were detected and their membrane lipids were characterized according to the ratio of the lipid methyl (—CH_3) to the methylene (—CH_2 —) groups.

Rank	Organism_GC	Coverage	DNA fraction	GC content	Scaffold_Gene (Sub9)	Scaffold_Gene (Sub27)	Gene identity
1	WWE3 RAAC2-related	566	0.058	0.429	scaffold_3564_6		ribosomal protein S3
2	WWE3 RAAC2-related	351	0.036	0.422		scaffold_6717_2	ribosomal protein L27
3	OP11 novel	270	0.028	0.454		scaffold_0_50	
4	OP11 ACD13-related	252	0.026	0.466	scaffold_231_6		ribosomal protein S3
5	OD1 novel	198	0.020	0.421	scaffold_0_137	scaffold_13_11	ribosomal protein S3
6	WWE3	158	0.016	0.425	scaffold_9_17		ribosomal protein S3
7	DPANN archaeon	135	0.014	0.353	scaffold_37_29		ribosomal protein S3P
8	OD1 RAAC4-related	90	0.009	0.527	scaffold_26_47		ribosomal protein L3
9	OD1 novel	90	0.009	0.443	scaffold_28_41		ribosomal protein S3
10	WWE3 RAAC2-related	90	0.009	0.430	scaffold_24_25		ribosomal protein S15
11	OD1 novel	72	0.007	0.399	scaffold_57_9		ribosomal protein S3
12	OP11 ACD13-related	63	0.006	0.411	scaffold_112_16		ribosomal protein S3
13	OP11_42	63	0.006	0.442	scaffold_67_8		ribosomal protein S3
14	OP11_41	63	0.006	0.423	scaffold_257_18		ribosomal protein S3
15	DPANN archaeon	54	0.006	0.357	scaffold_472_9		ribosomal protein S3AE

Supplementary Table 3. Community composition prior to acetate addition (GWA1) based on binning and analysis of assembled metagenomic data ($1/9^{th}$ and $1/27^{th}$ of the data). Sequences for marker genes are provided in **Supplementary Table 4**. Relative to the GWB1 post-acetate addition sample, GWA1 has more even community composition and lower representation of archaea and the Spirochaete.

gwb1_sub10_scaffold_461_15 rpS3
MGQKINPIGYRLGISRDWQSRWYAPASSYADVAHEDIKIRKYLKKKLDMAGLKEIDIERT
ENDISITVRVSKPGVVIGRGGTGVEEIEKEIKKLTKAKVKITAEEIKSPEIEAQLVGDYI
ARQIKRRMPYRRVVKFALSGAMDKGAKGIKIRLSGVLSGSNTISRSEQYTLGSIPLQTLR
ADIDYAQIDCHLLYGTIGIKVWIYKGEKTI*

gwb1_sub50_scaffold_8622_1 (incomplete) rpS3
IAFAVLVVVGNNNGRVGVGYGKAPDVATSINKAVSKAKKSMVEIKLNGSTISHEVIAKYE
SAKVFLKPAPKGTGVIAGGAVRPVLELAGIRDISAKMIGANNKISNVRCTLKALKKLKG*

gwb1_sub10_scaffold_28347_2 (version of gwb1_sub50_scaffold_8622_1) rppS3
MITADVNTAKFDEIEEKVLEIKRVSKKTKGGNTIAFAVLVVXXXXXXXXXXXXXXKFDEI
EEKVLEIKRVSKKTKGGNTIAFAVLVVVGNNNGRVGVGYGKAPDVATSINKAVGKAKKSM
VEIKLNGSTISHEVIAKYESAKVFLKPAPKGTGVIAGGAVRPVLELAGIRDISAKMIGAN
NKISNVRCTLKALKKLKG*

gwb1_sub50_scaffold_1_151 rpS3
MGKKVNPTIFRTGYIFPTKSVWYSNFKNYAKFVLEDNSIRRFLEEKLKLAGITNIEIKRS
INTVDIFMYVSRPGVVIGRGGSSLEQLKKDIEKLLKIDLKVKNAIKINLHPMEIKNPELS
AGIIVDRLSNQLEHRYPFRRAANQAIEKVMAAGAKGVKIVFAGRIDGAEIARTEKFKQGR
IPTQTLRANIDYVEKPALTRSGYVGIKVWIYTGDIII*

gwb1_sub10_scaffold_500_25 rpS3
MGQKINPIGMRVGEFIPWKSRWFSEKGFKNHLIEDIKIRKALMEKLKLAGITSVEIERLP
KSMAVTMTVSRPGVVIGRGGTGIEDVKKYVLGIIREVRKEKIEDLKIDLRVNEIKNPELS
AHLVATRIASELERRMPHRRVVTKTIERVMASGAAGVKVVLGGRIGGAEISRVEKFQEGS
VPTQTLRENIDYAQVPALLKRGYVGVKVWIHKKEAD*

gwb1_sub10_scaffold_3_60 rpS3
MGHKIKPTAFRVGVIKDWTARWFPKGVSFKEYLEEDVMIRKIIADKIGAAGIDYVAIERF
GDSIRIHIKAAKPGLIIGRGGKGIEDLSKLLDKKVAAIRVKNGSKKLTGISMNIEELKRF
DVSAAVTGYNIARDLERRMPFRRILKKTLENLMQTREVKGAKIRVSGRLNGAEIARTEQL
AKGTIPLQTIRADIDYAEATAYTTYGTIGVKVWINKGEVFKDKEQEPVKRQPRHDERRPR
IERKETRN*

gwb1_sub10_scaffold_130_13 rpS3
MGKKVNPIAFRTGYIFPTKSVWFANFRSYAKFVLEDNKIRRFLEEKLKLAGITSIEIKRS
INSVDIFMNVSRPGVVIGRGGSSLEILKKDLEKLLRINPKAKNAVKINLHPLEIKNPDIN
SSIIADRLIGQLEHRYPFRRAANQAIEKVMAAGAKGVKLVFAGRIDGAEIARVEKFKQGR
IPTQTLRANIDYIEKRALTRSGYVGIKVWIYTGDIIS*

gwb1_sub10_scaffold_110_21 rpS3
MGQKVNPTGFRIGTFLPWKSRWFADNNSFKQFLIEDIRIRKELTKKLKLAGITGVEIERS
PKSIVVTITVSRPGVVIGRGGTGIEDVKKYILGIMNDVRKKTVKDLKIDIRINEVKSPEL
SAYLVAERIVSEMERRIPHRRAVQKAIERVMASGALGIKVVLAGRINGAEISRVEKFHQG
SVPSQTLRENIDYAQVPALLKRGYVGVKVWIHKKAEA*

gwb1_sub10_scaffold_40_152 rpS3
MGNKIHPTNFRIGVIYNWKSRWLNRKFFKHFLEEDLRIREALAKQYNRSGLGEVEIERSG
DNATIIVNTAKPGLIIGRGGAGLTELKKKLEDIVKKLRLKGKYSTGKWELKLVVSEMKKP
ESEAKIVAQNIAADLEKRFPFRRAMKSALEKVTAQKDVLGVKINLAGRLNGSEMARREWL
SKGKVPLQTLRANIDYAQEEALTTYGKIGVKVWIYKGEIFEENK*

gwb1_sub10_scaffold_9_177 rpS3
MGQKINPVGYRLGISRDWQSRWYAPASSYADVAHEDIRIRKYLKKKLDMAGLKEIDIERT
ENDISITVRVSKPGVVIGRGGTGVEEIEKDIKKLTKAKVKITAEEVKSPEIEAQLVGDYI
ARQIKRRMPYRRVVKFALSGAMDKGAKGIKIRLSGVLSGSNTIARSEQYTLGSIPLQTLR
ADIDYAQIDCHLLYGTIGIKVWIYKGEKTI*

gwb1_sub10_scaffold_661_8 / gwb1_sub50_scaffold_856_7 rpS3
MGQKINPIGMRIGSFLPWKSRWFSEDRTFKKYLIEDIKIRGALFEKLKLAGITDVEIERL
PKSMVIYLTVSRPGVVIGRGGTGIEDVKRFIVKMLSVTKGKQTHDLKIDLKVGEVKNPEL
SAYLVAGRIVGELERRMPHRRVINRAMERVMASGARGIKVVLSGRIEGAEISRVEKYHMG
SVPTQTLRENIDYAQVPALGKRGYVGVKVWIHRKEEE*

gwb1_sub10_scaffold_61_58 rpS3
MGQKVNPIGLRIGINKDWSSRWFVDPRDYAKTLHEDLKIRARMYQLPEAKSADISEIEII
RHPQKVTIVIHTAKPGVLIGQKGANIERIGLELQKFASKKINIKIKEIKRVETNSMIVAM
NVCRQLEGRGSFRRTMKMAVSSAMKGGVQGCKVRMSGRLGGADMSRTEEYKEGRVPLHTL
RADIDYGFWEALTTYGKIGVKVWICKSDNALVGDKKEDAGLMPGRPRRDGGDRPERSERS
DRGGSNDRGAPRGPRTGGNRGGPRPTQARS*

gwb1_sub10_scaffold_67_52 rpS3
MGHKISPISFRVGIQRDWASRWFGTSKYIPLLKDDVAVRDYLEKKLKGMSVDRVEIERGT
DLLNVLIFSSRPGLLIGRGGAGIEDLKKALVKLLRKKVSIRIEIQEIKNPESSAAIMAES
IVDQVEKRIPFRRVMKQTLFKIMASRGVKGAKILMGGRLDGAEIARTEHLEEGSLPLSSL
RAEIDYAKKTARTTYGAIGIKVWIYKGLNFDKK*

gwb1_sub10_scaffold_484_24 rpS3
MGQKVHPRIFRIGIIYTWKSRWFSARGYREQLAVDVKLRRWLKKKLKGASVAGIEIERGP
SSLTVNILTAKPGVVIGRGGAQVEELKKEIKQKFLKSTDSLQLNIQEVTSPTLSAEIVAE
NMVMELEKRMPFRRVMKQAIDQSMKAGALGVKVEVGGRLNGAEIARTEKLIVGKVPLHTL
RADVDYSRSAAQTTYGVIGVKVWINKGEVLAPAEPNIMGNNQPAKSRKR*

gwb1_sub10_scaffold_4176_1 rpS3
TPSAYDILGVSSLDGDVQTPSTPALNVVDVTVVVEDSPVVEEEATVVEEEATVVEEEVTV
VVEEVTVVVEDSPVVEVSDEQLLMAEIERLLVERDELYGTINELRAENLELAEQAGEAEY
FIKRVAELEIEITRLEGIIVENEAALVAADTRLAEIEAQLVGDYLARQIKRRMPYRRVVK
FALSGAMDKGAKGIKIRLSGVLSGSNTISRSEQYTLGSIPLQTLRAEIDYAQIDCHLLYG
TIGIKVWIYKGEKTI*

gwb1_sub10_scaffold_210_14 rpS3P
MIEREFIKEKTKYLKIKEYIDSVIGSAAGVGKIIVEKTPLGEKILVYAVRPGLIIGRGGT
MIQELTVTLRNKFKLDNPQIEVTELENPNLSAAVMSKKVAADLERFGSSRFKAIGYRTLM
QIMKSGALGAEIMISGRGVPGARAHSWRFPAGHMKKSGQVALEQIDHVKTGANLRSGTVG
IQVRIMHPDIYQPDIIKIIEGVHIPEEVQKEIDMLEAEEQKAQKLLLKNTEPKPKKKKAE
ATDQIVTEIQEVKPELGSSQQGSANKSSLGKPKKRAKPKLSEAELGPAKQDVSLGKSDTD
SEKTNTISQDEVSEEPNENATQEANNEEVIQDTEKSGQ*

gwb1_sub10_scaffold_1424_4 rpS3
MGQKINPVGMRVGEFIPWKSRWFSEKGFKDYLIEDIKIRKALMEKLKLAGITSVEIERLP
KSMAVTMTVSRPGVVIGRGGTGIEDVKKYVLGIIREVRKEKVKDLKIDLRVNEIKNPELS
AYLVATRIVAELERRMPHRRVVTKTIERVMASGAKGVKVVLGGRIGGAEISRVEKFQEGS
VPTQTLRENIDYAQVPALLKRGYVGVKVWIHKKEAD*

gwb1_sub10_scaffold_389_4 rpS3
MTHSVHPYAHRLGIIRDWQSRWFGARGKYQEYLKADVLLREYLTKRLRGYYVDGVDIERG
AESLRITVRTSRPGLVIGRSGEGATKLKSDIQERLRAIGAAIPREFKLDIDEVKNPEAHA
GIVGYMIAEGLEKRLPFRRVAKQAIEKSMAAKGVQGVKVVLSGRLGGADMGRRESFKAGQ
VPLTTLRADVDFAREKAYLSYGVIGIKVWIYRGEVFES*

gwa1sub9_scaffold_3564_6 rpS3
MGQKINPIGYRLGISRDWQSRWYAPASSYADVAHEDIKIRKYLKKKLDMAGLKEIDIERT
ENDISITVRVSKPGVVIGRGGTGVEEIEKEIKKLTKAKVKITAEEIKSPEIEAQLVGDYI
ARQIKRRMPYRRVVKFALSGAMDKGAKGIKIRLSGVLSGSNTISRSEQYTLGSIPLQTLR
ADIDYAQIDCHLLYGTIGIKVWIYKGEKTI*

gwa1sub27_scaffold_6717_2 partial rpL27
MAHKKAGGSKARQGGNVAGKRLGVKVFGGSLVKAGGIIIRQRGRTFASGKNTDMAKDFSI

FSKVNGIVKFSWLTKKKKKIEVVKSE*

gwa1sub27_scaffold_0_50 rpS3
MRVGEFIPWKSRWFSEKGFKNHLIEDIKIRKALMEKLKLAGITSVEIERLPKSMAVTMTV
SRPGVVIGRGGTGIEDVKKYVLGIIREVRKEKIEDLKIDLRVNEIKNPELSAHLVATRIA
SELERRMPHRRVVTKTIERVMASGAAGVKVVLGGRIGGAEISRVEKFQEGSVPTQTLREN
IDYAQVPALLKRGYVGVKVWIHKKEAD*

gwa1sub9_scaffold_231_6 rpS3
MGQKINPIGMRVGEFIPWKSRWFSEKGFKNHLIEDIKIRKALMEKLKLAGITSVEIERLP
KSMAVTMTVSRPGVVIGRGGTGIEDVKKYVLGIIREVRKEKIEDLKIDLRVNEIKNPELS
AHLVATRIASELERRMPHRRVVTKTIERVMASGAAGVKVVLGGRIGGAEISRVEKFQEGS
VPTQTLRENIDYAQVPALLKRGYVGVKVWIHKKEAD*

gwa1sub9_scaffold_0_137 rpS3
MGNKIHPTNFRIGVIYNWKSRWLNRKFFKHFLEEDLRIREALAKQYNRSGLGEVEIERSG
DNATIIVNTAKPGLIIGRGGAGLTELKKKLEDIVKKLRLKGKYSTGKWELKLVVSEMKKP
ESEAKIVAQNIAADLEKRFPFRRAMKSALEKVTAQKDVLGVKINLAGRLNGSEMARREWL
SKGKVPLQTLRANIDYAQEEALTTYGKIGVKVWIYKGEIFEENK*

gwa1sub27_scaffold_13_11 rpS3 (as above)
MGNKIHPTNFRIGVIYNWKSRWLNRKFFKHFLEEDLRIREALAKQYNRSGLGEVEIERSG
DNATIIVNTAKPGLIIGRGGAGLTELKKKLEDIVKKLRLKGKYSTGKWELKLVVSEMKKP
ESEAKIVAQNIAADLEKRFPFRRAMKSALEKVTAQKDVLGVKINLAGRLNGSEMARREWL
SKGKVPLQTLRANIDYAQEEALTTYGKIGVKVWIYKGEIFEENK*

gwa1sub9_scaffold_9_17 # 17430 # 18062 # 1 # ;gc_cont=0.425
MGQKINPVGYRLGISRDWQSRWYAPASSYADVAHEDIRIRKYLKKKLDMAGLKEIDIERT
ENDISITVRVSKPGVVIGRGGTGVEEIEKDIKKLTKAKVKITAEEVKSPEIEAQLVGDYI
ARQIKRRMPYRRVVKFALSGAMDKGAKGIKIRLSGVLSGSNTIARSEQYTLGSIPLQTLR
ADIDYAQIDCHLLYGTIGIKVWIYKGEKTI*

gwa1sub9_scaffold_37_29 rpS3P
MIEREFIKEKTKYLKIKEYIDSVIGSAAGVGKIIVEKTPLGEKILVYAVRPGLIIGRGGT
MIQELTVTLRNKFKLDNPQIEVTELENPNLSAAVMSKKVAADLERFGSSRFKAIGYRTLM
QIMKSGALGAEIMISGRGVPGARAHSWRFPAGHMKKSGQVALEQIDHVKTGANLRSGTVG
IQVRIMHPDIYQPDIIKIIEGVHIPEEVQKEIDMLEAEEQKAQKLLLKNTEPKPKKKKAE
ATDQIVTEIQEVKPELGSSQQGSANKSSLGKPKKRAKPKLSEAELGPAKQDVSLGKSDTD
SEKTNTISQDEVSEEPNENATQEANNEEVIQDTEKSGQ*

gwa1sub9_scaffold_26_47 rpS3
MKYLLATKIAMTELFDESGNAHGATVLKTGPLTVTQLRTKEKDGYAALQAGFVAAKEKSL
TKPLKGHLKASGGLFKHLKEFPLDGDMEVKVGDTVGLSQFAPGDTVTVRGISKGKGFQGV
VKRHGFHGGPRSHGQKHSEREPGSIGASGVQRVYKGVRMAGRMGGDAITVKNLTVLKVDA
ERSELYLKGAIPGRRGTVLTVIGR*

gwa1sub9_scaffold_28_41 rpS3
MTHTVHPYSHRLGILRDWKSRWFSPKGRYKYALRGDILIRQYLEKKLKGFFVALIEIERS
EKSLRIILSSSRPGMIIGRQGDGAELLRKEIKSLLLKHKLLEEKVDVRLDIKEIKSAETN
AAIVSQMIVEGLEKRLPFRRVMKQMLEKVMANRDVLGVRIFLAGRLGGATMSRTEDRKLG
RIPLQTLRADVDYALTKAVMPYGTIGVKVWIYKGDIFK*

gwa1sub9_scaffold_24_25 rpS3 partial
MALNKDKKASIIKTNRLHDADTGSPEVQIALLNEKITKLSAHLKAHKKDNHSRRGLLQMV
NKRRRLLSYLKKKDEERFTSVSEKLELSK*

gwa1sub9_scaffold_57_9 rpS3 partial
MAESIVDQVEKRIPFRRVMKQTLFKIMASRGVKGAKILMGGRLDGAEIARTEHLEEGSLP
LSSLRAEIDYAKKTARTTYGAIGIKVWIYKGLNFDKK*

gwa1sub9_scaffold_112_16 rpS3
MGQKVNPTGFRIGTFLPWKSRWFADNNSFKQFLIEDIRIRKELTKKLKLAGITGVEIERS
PKSIVVTITVSRPGVVIGRGGTGIEDVKKYILGIMNDVRKKTVKDLKIDIRINEVKSPEL
SAYLVAERIVSEMERRIPHRRAVQKAIERVMASGALGIKVVLAGRINGAEISRVEKFHQG
SVPSQTLRENIDYAQVPALLKRGYVGVKVWIHKKAEA*

gwa1sub9_scaffold_67_8 rpS3 partial
MGQKINPVGMRVGEFIPWKSRWFSEKGFKDYLIEDIKIRKALMEKLKLAGITSVEIERLP
KSMAVTMTVSRPGVVIGRGGTGIEDVKKYVLGIIREVRKEKVKDLKIDLRVNGQGCFGRK
NRRGRNFQG*

gwa1sub9_scaffold_257_18 rpS3 partial
MGQKINPVGFRMGGTAFWQSRWFADDKKYRQFIAEDMKIRNLLMKKLRPAGVARVEIERS
INKVKIIIFVARPGVLIGRGGTGLLDLKKLLMKQLDIKNENTLEVMPMDVKSPDLSAYLV
AQNVVEQLIRRLPAQRVMNQTIERVMRAGAKGVKVVLSGRIGGAEIARRERKAAGTMPLH
TLRQ

gwa1sub9_scaffold_472_9 rpS3
MAVAKKRKRFFDVDIPLIDKETHLQAFEIDELKGRMIKYDLTRMLKGKSMILISKIKVEN
DKAISVPREIRLMPYFLRRMVRKGTNYIEDSFIAQCIDTKVRIKPFLITRRKVSRAIRSA
LRQKTKTELENYLKNKNSEEIFSEILSGQIQKQFSLILKKIYPLSLFEIRVLKIEGKEKD*

Supplementary Table 4. Ribosomal protein sequences of marker genes used for the GWA1 community composition analysis (prior to acetate amendment).

Rank	Organism	GC content (average)	Coverage (Sub10)	DNA fraction	Coverage (Sub50)	Scaffold_Gene (Sub10)	Scaffold_Gene (Sub50)	Gene identity
1	WWE3 related to RAAC2	0.435	1210	0.070	1050	gwb1_sub10_scaffold_461_15		30S ribosomal protein S3
2	WWE3 related to RAAC2	0.433	750	0.043	600		gwb1_sub50_scaffold_8622_1	30S ribosomal protein S3
3	OP11 related to RAAC19	0.448	490	0.028	500		gwb1_sub50_scaffold_1_151	30S ribosomal protein S3
4	OP11 related to ACD13	0.469	419	0.024	420	gwb1_sub10_scaffold_500_25		30S ribosomal protein S3
5	OD1 related to ACD81	0.459	390	0.022	380	gwb1_sub10_scaffold_3_60		30S ribosomal protein S3
6	OP11 related to ACD61	0.455	316	0.018	316	gwb1_sub10_scaffold_130_13		30S ribosomal protein S3
7	OP11 related to ACD13	0.408	193	0.011	224	gwb1_sub10_scaffold_110_21		30S ribosomal protein S3
8	OD1 novel	0.431	187	0.011	N/A	gwb1_sub10_scaffold_40_152		30S ribosomal protein S3
9	WWE3 related to RAAC2	0.429	183	0.011	N/A	gwb1_sub10_scaffold_9_177		30S ribosomal protein S3
10	OP11 novel	0.401	161	0.009	N/A		gwb1_sub50_scaffold_856_7	30S ribosomal protein S3
11	Spirochaetales	0.629	124	0.020	N/A	gwb1_sub10_scaffold_61_58		30S ribosomal protein S3
12	OD1 novel	0.395	117	0.007	N/A	gwb1_sub10_scaffold_67_52		30S ribosomal protein S3
13	OD1 novel	0.433	105	0.006	N/A	gwb1_sub10_scaffold_484_24		30S ribosomal protein S3
14	WWE3 related to RAAC2	0.427	85	0.005	N/A	gwb1_sub10_scaffold_4176_1		30S ribosomal protein S3
15	DPANN archaeon related to AR3	0.370	84	0.005	N/A	gwb1_sub10_scaffold_210_14		30S ribosomal protein S3P
16	OP11 novel	0.468	80	0.005	N/A	gwb1_sub10_scaffold_1424_4		30S ribosomal protein S3
17	OD1 related to RAAC4	0.528	70	0.004	N/A	gwb1_sub10_scaffold_389_4		30S ribosomal protein S3

Supplementary Table 5. Representation of the more abundant organisms in the GWB1 sample (after acetate injection). Coverage statistics are listed for assemblies for 10% (Sub10) and 2% (Sub50) of the data. Bin references use ribosomal protein S3. Sequences are deposited under accession number KC999117-KC999376. Abbreviation: N/A, not available.

	Cell		Cytoplasmic space		Cell envelope
	major axes (nm)	minor axes (nm)	major axes (nm)	minor axes (nm)	thickness (nm)
Mean	333	240	284	198	23
Median	318	241	273	198	21
Minimum	195	149	180	125	5
Maximum	552	356	485	298	43
Standard deviation	75.48	44.95	68.35	38.08	6.37
Standard error of the mean	8.28	4.93	7.50	4.18	0.70

Supplementary Table 6. Size distribution of WWE3-OP11-OD1 ultra-small bacteria. The statistics are based on 83 two-dimensional high resolution cryogenic micrographs. Only three cells had diameters of > 500 nm.

a

	cryo-TEM	Clone library	EMIRGE
Sample size	63 cells	24 OTUs	36 OTUs
Ultra-small cells (%)	74.60	75.50	78.96
Spirochaetes (%)	4.76	21.10	13.96
Other bacteria (%)	20.63	3.50	6.76

b

	cryo-TEM to clone library	cryo-TEM to EMIRGE
z-score	0.4899	0.0865
p-value	0.6244	0.9283
	not significant at $p < 0.05$	not significant at $p < 0.05$

Supplementary Table 7. Linkage of molecular data that provides information about WWE3, OP11, and OD1 abundances in sample GWB1 and imaged cells. To link the molecular data on WWE3, OP11, and OD1 and the ultra-small cells observed by cryo-TEM, a z-test for 2 population proportions, two tailed, was performed. For cryo-TEM data, grids were surveyed objectively, every cell was categorised as ultra-small cells, Spirochaetes or other bacteria. **a**) The cryo-TEM data of the ultra-small cells were then compared to the results found by clone library and EMIRGE for the 0.1 μm filter (**Fig. 1a**). **b**) For ultra-small cells no significant differences were found, supporting the inference that the majority of images cells were CP radiation bacteria.

Target	Probe	*E. coli* position	Probe sequence	Optimal formamide concentration (%)
WWE3	WWE3-1418	1418-1438	CCACCGTGAGCCCACTAATG	50
OP11	OP11-828	828-850	CACTCCCCTTACGGGAAATAGC	50
OD1	OD1-1250	1250-1270	CTTACGGCTTGGCAACCCTT	50

Supplementary Table 8. CARD-FISH probes used to analyse ultra-small bacteria in the WWE3-OP11-OD1 radiation present in ~0.2 μm filtered groundwater after acetate amendment at the Rifle field site. Due to the high diversity within these CP clades, it was impossible to design probes targeting the entire clades. Hence, we designed probes that specifically targeted the WWE3, OP11 and OD1 cells that were dominant in our samples, providing the highest probability for successful labelling.

Supplementary Note 1.

Synchrotron infrared (SIR) spectromicroscopy for determination of community composition of ultra-small cells

A combined Principal Component Analysis and Linear Discriminant Analysis (PCA-LDA) was performed on the spectra of the ultra-small cells and compared to spectra from model bacteria: *Escherichia coli* and *Bacillus atrophaeus*, and archaea: *Methanopyrus kandleri* and *Sulfolobus solfataricus*. The spectra of the model bacteria and archaea were from our validated database library. Each group contained the same number of spectra, n = 800. The data were normalized by standard deviation (STD normalization) on the 1550 cm^{-1} band absorption intensity and the spectral region of interest selected for the analysis was 2800-3000 cm^{-1}.

The PCA-LDA factor 2 separates the bacteria from archaea (**Supplementary Fig. 4a)**. The ultra-small cell dataset is located in the bacteria hemispace, together with *E.coli* and *B. atrophaeus*. PCA-LDA factor 2 presents two sharp negative spectral features at 2920 cm^{-1} and 2850 cm^{-1} (arrows in **Supplementary Fig. 4b**), assigned to antisymmetric and symmetric C–H stretching vibrations of —CH_2– groups[1], respectively. Considering the position of the ultra-small cells in the PCA-LDA quarter, it is possible to speculate that the ultra–small cell membrane is composed of lipids characterized by little branching and/or little saturation.

Supplementary Note 2.

2D and 3D cryo-TEM

In low-dose cryo-TEM, the first step is to carry out an exhaustive survey of the sample for selection of cells in optimal relationship to the grid bars to enable 3D data acquisition (these cells must be imaged at very low dose prior to 3D imaging (low signal to noise ratio), and

thus images are not publication quality). From this analysis, we determined that the two acetate-stimulated groundwater samples were highly dominated by very small cells. Beyond this, we extensively surveyed many samples and recorded over 100 high quality (non-survey; high signal to noise ratio), non-tomographic, 2D images of ultra-small cells. Additionally 13 tomographic tilt series were acquired under low dose conditions and reconstructed.

Supplementary Note 3.

Morphotypes

We selected 78 highest quality (based on transparency, absence of dirt, artefacts, and overall image quality) 2D images from the complete data set, obtained with different magnifications, to describe different morphotypes of the ultra-small cells. Additionally, 13 tomograms were analysed to determine cell morphology. **Supplementary Fig. 6** represents the 10 most common morphotypes of ultra-small cells found in our sample. Following categories were chosen: **a**: round cell, a surface layer, few to many radiating pili-like structures. **b**: elongate cell with a surface layer **c**: oval-shaped cell, surface layer. **d**: egg-shaped cell, surface layer and polar pili-like structures. **e**: egg-shaped cell, surface layer, many radiating pili-like structures. **f**: elongate thin cell, polar pili-like structures. **g**: irregular shaped cell, polar pili-like structures. **h**: rounded oval-shaped cell. **i**: irregular round cell (sometimes one polar pili-like structure). **j**: round cell, lots of radiating pili-like structures (sometimes no pili-like structures).

Supplementary Note 4.

Surface Protein Layer (S-Layer)

The cell envelope has a novel architecture, with a remarkable and distinct S- layer, and several appendages (**Supplementary Fig. 9**, **10** and **11**). A striking pattern of parallel lines is

seen in many 2D projections (**Supplementary Fig. 9b**). Closer inspection reveals they are part of the S-layer structure on the cell surface and have a periodicity of 3.3 nm. The basic repeating motif of the S-layer is hexagonal and is made of small subunits approximately 1 nm in diameter spaced with centers of mass of approximately ~3 nm from each other (**Supplementary Fig. 9a**). The distance from the centre of one hexagon to the centre of another one is 13 nm. The thickness of the S-layer (not including the connectors) is ~ 13 to 14 nm. Since these microorganisms are oblate and often near spherical, it is interesting the selected tiling pattern should be hexagonal. We have been unable to visualize what types of defects are most common in order to achieve maximum coverage.

Supplementary Note 5.

Subtomographic Averaged Reconstructions

For subtomographic averaged reconstructions, three whole cell reconstructions were surveyed with Imod and the locations of 1,167 S-layer lattice units from three low defocus tomographic reconstructions (382, 410, 375). The chosen microorganisms belong to morphtotype categories A, D and E. Locations were manually chosen and stored in segmented models. Cubical subvolumes (64 voxels by side, 0.56 nm^3), with assigned normal pointing outwards from the cell surface, were cropped. The side of the cubical volume was about twice the lattice constant and contained a centred repeating unit. The centre of each repeating unit in the subvolume was aligned for averaging 3D S-layer lattices. This process allowed us to compute the centre of mass of each cropped sub-volume and to use a cell surface normal at each point for rotational alignment of all subvolumes. In whole cell data the normal defines the outside of the bacterium and allows the merging of data from different cryo-tomograms. A first model was obtained computing the iterative alignment and averaging of 382 subvolumes cropped from one data set acquired with defocus value of ~6 μm ± 0.25

μm. For the final refinement of the subtomographic averaged reconstructions shown here 785 subvolumes were used; these were cropped from 2 data sets acquired using defocus value of ~4 μm ± 0.25 μm (target value 3.6 μm). Maximum Likelihood iterative classification and average structure determination resulted in only one basic underlying structure for all 1,200 selected units from 3 cells. For subtomographic averaged reconstructions, three microorganisms were randomly chosen at low dose diffraction defocus.

Alignment and classification of the boxed sub-volumes were computed with the various utilities within the "X-Window-based Microscopy Image Processing Package", or Xmipp package, (http://xmipp.cnb.csic.es/twiki/bin/view/Xmipp/WebHome). Several clustering or classification strategies using different algorithms were used in order to validate the results across conceptually different methodologies.

Supplementary Note 6

Linking cryo-TEM to Catalysed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) signal

Our aim was to analyse intact ultra-small bacteria by cryo-TEM first, to obtain information on morphology and ultrastructure. We hoped to label and identify the exact same cell as previously described[2]. However, linking cryo-TEM images to CARD-FISH signal detected via optical microscopy is very challenging, even for normal-sized bacterial and archaeal cells. Damage to rRNA caused by the electron beam decreases or eliminates CARD-FISH labelling efficiency. Therefore, we also applied CARD-FISH first, detected CARD-FISH signals by optical microscopy and then carried out transmission electron microscopy. A drawback of this approach is that labelling protocols such as FISH destroy the ultrastructure of small microorganisms beyond recognition[3].

We designed CARD-FISH probes to target rRNA sequences recovered via the clone

library analysis (**Supplementary Table 8**). The probes were designed to specifically target the most abundant bacteria from the OD1, OP11 and WWE3 lineages in our samples (not for sequences in public databases, such as SILVA, which may be only distantly related).

As positive controls, the rRNA sequences were engineered onto plasmids carried by *E. coli* cells[4]. To establish stringency concentrations of formamide in the hybridization buffer were changed at fixed incubation temperature. At a formamide concentration of 50% the best fluorescence signal of the target cells was achieved. A 50 % formamide concentration also prevented non-specific binding when using two different probes in one sample.

Despite successful labelling of the positive controls, no fluorescence could be detected from cells binding the WWE3, OP11 or OD1 CARD-FISH probes, regardless of whether TEM preceded or following the FISH experiment. Many very small, fluorescent dots were detected in FISH experiments by optical microscopy when CARD-FISH was performed prior to TEM. However, across many attempts, cryo-TEM characterization of the exact areas showing probe fluorescence did not detect cells. Either the cells were lost during specimen transfer or fluorescent signal was due to non-specific probe binding.

One complication with the FISH experiments might have been that the cell wall composition of these organisms may be different to that of other bacteria for which probe methods have been developed. For this reason, our efforts to label ultra-small cells by FISH included attempts to permeable the cells in a variety of ways, including with a lysozyme solution, SDS and hydrochloric acid. However, the cell envelope may have presented a boundary to the penetration of the probes into the cells. Other complications with FISH labelling may be the very low ribosome content and tight packing of the ribosomes, which may have inhibited access of the probe to the rRNA.

Supplementary References

1. Snyder RG, Hsu SL, Krimm S. Vibrational-spectra in C-H stretching region and structure of polymethylene chain. *Spectrochim Acta A* **34**, 395-406 (1978).
2. Knierim B*, et al.* Correlative microscopy for phylogenetic and ultrastructural characterization of microbial communities. *Environ Microbiol Rep* **4**, 36-41 (2011).
3. Comolli LR, Baker BJ, Downing KH, Siegerist CE, Banfield JF. Three-dimensional analysis of the structure and ecology of a novel, ultra-small archaeon. *ISME J* **3**, 159-167 (2009).
4. Schramm A, Fuchs BM, Nielsen JL, Tonolla M, Stahl DA. Fluorescence *in situ* hybridization of 16S rRNA gene clones (Clone-FISH) for probe validation and screening of clone libraries. *Environ Microbiol* **4**, 713-720 (2002).
5. Wrighton KC*, et al.* Fermentation, Hydrogen, and Sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**, 1661-1665 (2012).